

Machine and Computer-assisted Interpreting: Innovations in and Implications for Interpreting Practice, Pedagogy and Research

Xinchao Lu

Beijing Foreign Studies University

luxinchao@bfsu.edu.cn

<https://orcid.org/0009-0000-4195-9101>

Claudio Fantinuoli

Mainz University

fantinuoli@uni-mainz.de

<https://orcid.org/0000-0003-1312-0741>

Abstract

Transformative advancements in technologies such as Automatic Speech Recognition, Natural Language Processing, Deep Learning, and Generative AI have significantly accelerated the evolution of both Machine Interpreting (MI) and Computer-Assisted Interpreting (CAI), fundamentally reshaping the interpreting ecosystem. MI has progressed from being based on statistical machine-translation (MT) models to being augmented by large-language and multimodal models, while undergoing a transition from cascade to end-to-end systems. These developments have markedly enhanced MI's capacity to manage increasingly complex and diverse domains, linguistic features, and operational contexts. CAI, originally designed to streamline the preparatory processes for interpreting, has evolved to incorporate real-time functionalities that are integrated into the interpreting workflow, enabling the effective management of complex terminology and other problem-triggers. This introduction begins by providing a comprehensive overview of the evolution, current state, and future directions of MI and CAI, followed by an introduction to the seven contributions featured in this special issue. These studies encompass a diverse range of topics, including MI quality evaluation, a comparison of human-machine interpreting products and process, CAI tool assessment in remote simultaneous interpreting (SI), the implications of automatic speech recognition (ASR) for consecutive interpreting, the effect of multimodal inputs, and the user-machine interaction patterns with live captioning in SI. This introduction, along with the seven contributions in this issue, aims to advance the growing body of knowledge on the

transformative impact of MI and CAI in reshaping interpreting practice, pedagogy, and research.

Keywords: machine Interpreting (MI), computer-assisted interpreting (CAI), interpreting quality, interpreting process, human–machine collaboration

1. Introduction

Over the past century, information and communication technologies (ICTs) have reshaped interpreting practice: the introduction of wired systems for speech transmission giving birth to simultaneous interpreting (SI) in the early 1920s; the boom of the World Wide Web in the 1990s having radically changed interpreters' acquisition of knowledge and terminology; a technological turn since the 1980s that manifested itself in the development of machine interpreting, computer-assisted interpreting (CAI) and remote interpreting (Fantinuoli, 2018; Tripepi Winteringham, 2010).

Machine interpreting (MI, also termed “speech translation”, “speech-to-text/speech translation”, “spoken language translation”, “simultaneous machine translation/interpreting”, or “automatic interpreting”) refers to the practice, process, or product of real-time automatic or automated speech translation by a computerized cascade system composed of subsystems of automatic speech recognition (ASR), machine translation (MT), speech synthesis, and subtitling or, more recently, by an end-to-end system without the conventional subsystems. Here, the term is understood broadly to encompass any translation from spoken input to written and/or spoken output marked by immediacy, its real-time use to bridge language barriers in a specific moment and setting, and its clear overlap with the activity performed by interpreters and re-speakers. The term “machine interpreting” was first mentioned by Salevsky (1993) as one of three categories – human, machine and computer-aided interpreting – when she was outlining areas of “special theories within interpreting studies”. MI seems to be a term used only by the translation and interpreting community, whereas “speech” plus “translation” appears in nearly all publications generated by the computational community.

Between the 1980s and the 1990s, MI was applied in areas such as reservations and scheduling, and from the early 2000s its application expanded to include everyday travel conversations and speeches (Nakamura, 2009). During the next two decades, remarkable advancements in ASR, natural language processing (NLP), deep learning (DL), neural machine translation (NMT), and artificial intelligence (AI) have given a major boost to the development of MI systems. These advancements have enhanced system robustness in handling increasingly complex and diverse source-language features and environments, while also expanding their applications across domains, modes, and scenarios through improved acceptability, affordability,

portability, and usability. More recently, the rise of Generative AI, particularly Large Language Models (LLMs), has catalysed transformative advancements in this field.

Computer-assisted interpreting (CAI) refers to the use of digital tools – such as desktop software, mobile applications, and, more recently, AI-driven systems, that support human interpreters at various stages of their work – from preparation to performance and post-task analysis. First introduced in the 2000s to support the preparatory phase of interpreting assignments, CAI tools have undergone substantial development over the past decade and are now capable of assisting interpreters in real-time. These tools facilitate terminology management, automate document analysis, and support knowledge acquisition by providing interpreters with rapid access to domain-specific content. The term *computer-assisted interpreting*, now commonly used, was first introduced by Fantinuoli (2017), building on the earlier term *computer-aided interpreting* introduced by Salevsky (1993) and marking the beginning of a more systematic approach to the research and integration of digital tools in interpreting. Since then, CAI has evolved into a mainstream technological category, with professional interpreters, some of whom work in national and international organizations, incorporating dedicated CAI platforms into their workflows.

As mentioned above, the early generation of CAI tools focused primarily on preparation, streamlining processes such as glossary compilation, terminology extraction from parallel corpora, and flashcard-based memorization. The aim was to reduce the cognitive load during interpreting by maximizing anticipatory information-processing in advance of an assignment. More recently, however, technological advancements, particularly in ASR and NMT, have introduced real-time functionalities that integrate directly into the live interpreting workflow. These developments have attracted considerable academic and professional interest and empirical studies are increasingly demonstrating that such tools are able to enhance interpreters' performance, especially in the management of problem-triggers such as names, numbers, and specialized terminology.

In the light of these transformative developments, the publication of a special issue of *Linguistica Antverpiensia, New Series: Themes in Translation Studies (LANS-TTS)*, whose theme for 2025 is Machine and Computer-assisted Interpreting, is both timely and essential. This issue aims to examine and explore critically the emerging topics in this rapidly evolving field. The remainder of this introduction is structured as follows: section 2 provides a comprehensive overview of recent research advancements in MI while section 3 focuses on developments in CAI. Section 4 introduces the seven studies featured in this special issue and the concluding section identifies areas for further consideration and exploration.

2. Evolution of and state of the art in machine interpreting

Research on MI has been undertaken by two distinct groups: the computational research community and the translation and interpreting research community. The former has been the primary force behind the innovations and advancements in developing MI systems, whereas the latter has focused largely on system evaluation or human–machine performance comparison. The current state of the art is characterized by a lack of interaction and exchange between the two communities (Pöchhacker, 2024). In the following sections, we present the main contributions of the two research communities to this topic.

2.1 Contributions of the computational research community

2.1.1 System and architecture development

MI systems before 2016 were generally cascade systems that developed with the goal of creating telephones with the integrated ability to translate dialogues (Luperfoy, 1996) or systems to translate lectures (Cho et al., 2013; Fügen et al., 2007). To overcome challenges such as error accumulation, high computational demands, and inference delays that are common in cascade systems, Jia et al. (2019) proposed an attention-based sequence-to-sequence neural network capable of performing direct speech-to-speech translation. End-to-end systems integrate speech recognition and translation in a single encoder–decoder neural network. Such end-to-end systems have achieved comparable results to or even better results than their cascade equivalents (see Fang & Yeng, 2023; Ma et al., 2024) and, more recently, LLM-powered high-performance end-to-end systems have appeared (see Chen et al., 2024; Cheng et al., 2024).

2.1.2 Corpus construction

For the cascade systems, large language datasets are required to train the various models that comprise the pipeline, including ASR, MT and speech synthesis. High-quality corpora across languages, domains, topics, and other areas are also crucial to a system’s learning and adapting to improve its performance and accuracy. Research on corpus construction has ranged from preparing corpora for the translation of travel dialogues (Kikui et al., 2006) to building a corpus for SI lectures (Murata et al., 2010) and incorporating interpreter data or annotations into machine-learning systems (Cheng et al., 2024; Shimizu et al., 2013). End-to-end systems face a major challenge due to the lack of sufficient speech–translation paired corpora for neural network training. However, data augmentation, which creates new training data from existing datasets, may offer a solution (see Lam et al., 2023; Mi et al., 2022).

2.1.3 Model development

MI systems have evolved from statistical to neural models and, more recently, to LLM-based models. Cascade systems such as MASTOR (Gao et al., 2006) generally require multiple models, ranging from source-acoustic, source-language and translation models to target-language and speech-synthesis models. Novel decoding algorithms (e.g., Cho & Esipova, 2016) enabled NMT to start before receiving a full source sentence, which paved the way for neural simultaneous translation (ST) systems. Besides cascade methods, attention-based (Jia et al., 2019) and LLM-based (Dong et al., 2023) models enable end-to-end translation while preserving voice and speaking style. Advances in Big Data and LLMs enabled the development of an effective speech translation model based on pre-trained LLMs (Huang et al., 2023). LLM-ST integrates LLMs with a speech encoder and multi-task tuning for precise timestamped transcriptions and translations of long audio. ComSpeech, a composite speech-to-speech translation model, connects pre-trained speech-to-text and text-to-speech models (Fang et al., 2024; SEAMLESS Communication Team, 2025).

2.1.4 Quality evaluation

MI quality is evaluated through source–target comparison, target–reference comparison, or user reception. This can be performed either manually (e.g., by interpreters themselves) or automatically using metrics such as BLEU, CHRF, COMET, METEOR, TER, and WER. Automatic evaluation, while currently less reliable than its human equivalent, offers greater efficiency and cost-effectiveness. For cascade systems, both component and overall system quality need to be evaluated (Hamon et al., 2009; Le et al., 2018). Studies have also explored user experiences (Sakamoto et al., 2013) and behavioural changes (Shin et al., 2013). Despite advancements in end-to-end systems, quality is still primarily evaluated using text-based metrics, a function that relies on automatic speech transcription. A notable exception is BLASER, introduced by Chen et al. (2022), which is a text-free metric designed for use in end-to-end systems and which leverages a multilingual multimodal encoder to embed speech input, translation output and references into a shared space for direct quality scoring.

Systems, corpora, models, and quality are the key research topics in MI, with studies offering either holistic overviews or focused analyses of specific issues or processes – for example, speech recognition of tonal speech signals (Dua et al., 2022); sentence segmentation (Siahbani et al., 2018); punctuation handling (Wang et al., 2016); decoding algorithms (Cho & Esipova, 2016); incremental processing, anticipation and latency (Grissom II et al., 2014; Ma et al., 2019); processing speech disfluencies (Mujadia et al., 2025); contextual information capturing and exploitation (Sridhar et al., 2013); facial expression-based affective speech translation (Székely et al., 2014); multimodal speech translation (Wu et al., 2019); a holistic cascade system for expressive speech-to-speech translation (Huang et al., 2023); direct speech

Lu, X., & Fantinuoli, C. (2025). Machine and computer-assisted interpreting: Innovations in and implications for Interpreting practice, Pedagogy and Research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 1–22.

translation for automatic subtitling (Papi et al., 2023), and end-to-end simultaneous speech-to-any translation (Ma et al., 2024).

2.2 Contributions of the translation and interpreting research community

In contrast to the computational community, where MI has been extensively studied, albeit under the expression “speech-to-speech translation system”, since the 1990s, it has remained underexplored in the translation and interpreting community. A search of the Benjamins Translation Studies Bibliography (TSB), for instance, yields fewer than ten entries of publications on MI from the late 1990s to the early 2020s (Pöchhacker, 2024). And although they published in the journal *Interpreting*, Jekat and Klein (1996) and Luperfoy (1996) have remained the works of the computational community. Research on interpreting in the journal *Machine Translation* is limited: in its special issue on spoken-language translation, for example, only one article dealt with discourse particles and discourse functions (Stede & Schmitz, 2000) while the rest were authored by computer scientists and technologists.

Translation and interpreting scholars have focused more on CAI than on MI, as evidenced in this current issue of *Linguistica Antverpiensia*. Research has focused on evaluating MI quality, comparing it with human interpreting with respect to competence and process. Unlike the computational community’s reliance on text-based metrics, translation and interpreting scholars prioritize information fidelity and user perception. This latter approach suits most interpreting contexts, emphasizing effective information transfer over linguistic equivalence amid diverse linguistic and situational features.

The master’s studies by Wonisch (2017) and Glocknitzer (2020) (both cited in Pöchhacker, 2024) comprised an evaluation of MI apps in dialogue settings. Fantinuoli and Prandi (2021) adopted a user-centric and communication-oriented methodology to evaluate the performances of a commercial speech-translation engine and professional simultaneous interpreters, finding that human beings excelled in the area of intelligibility whereas the machine slightly outperformed them in the area of informativeness. Lu (2022) compared the fidelity, idiomaticity, and usability of Chinese–English SI by professionals and a machine, revealing mixed-quality results across two distinct contexts: prepared opening remarks at a political forum and an improvised presentation at a business forum. Lu (2023) compared the processes, competencies, and quality of human and machine SI, highlighting their complementarity and advocating a hybrid model that integrates machine-assisted human interpreting and human-assisted MI. Liu & Liang (2024) identified significant linguistic differences between human interpreting and MI across lexical, syntactic, and cohesive features. The human beings excelled in audience-oriented communication but faced cognitive limits, whereas the machines reduced interpreting to word-to-word information transfer.

Whereas both human interpreters and MI exhibit distinct strengths, the overall performance of current cascade systems remains inferior to that of professional interpreters. Nevertheless, the emergence of end-to-end systems, LLMs, and the iterative nature of machine systems are gradually reshaping this dynamic – as demonstrated in the subsequent sections.

2.3 Cooperation between the two communities

Based on an analysis of cascade system performance, Lu (2022) proposed prioritizing the refinement of human interpreters' processing advantages to enable machines to emulate professional interpreters' cognitive mechanisms more effectively. Cheng et al. (2024) introduced CLASI, an end-to-end LLM-based speech-to-text system designed for high-quality human-like simultaneous translation. Its performance draws on professional interpreters' expertise in chunking, advance preparation, context-based processing, disfluency handling, and message-based evaluation.

CLASI adopted a robust read-write policy for simultaneous speech translation derived from professional interpreters' annotations of real-world speech. These annotations captured read-write timing for segmentation, that is, semantic chunking through syntactic boundaries (e.g., pauses, commas, conjunctions) and contextual meaning. Using a data-driven learning process, CLASI was trained to emulate professional interpreters. This strategy effectively balanced translation quality and latency while preventing output rewriting – a common limitation in other systems. Inspired by human interpreters' preparatory processes, the authors introduced a Multi-Modal Retrieval Augmented Generation (MM-RAG) framework. A multi-modal retriever extracts knowledge from an external database based on speech input. The retrieved information and the contextual memory are appended to the LLM prompt to enhance translation via in-context learning. In addition, the authors introduced a Multi-task Supervised Fine-tuning Training method, leveraging human-annotated data to refine their model and enhance its robustness and translation quality. The resultant high-quality data align the model with professional interpreters' segmentation strategies and improve its handling of speech disfluencies, facilitating more effective communication in real-world scenarios. The authors collaborated with professional interpreters to create a new evaluation metric – Valid Information Proportion (VIP) – which measures the percentage of accurately conveyed information. Human evaluations on diverse, challenging, and long speech datasets demonstrated the significant superiority of their approach over existing systems. For instance, CLASI achieved a VIP score of 81.3% in Chinese-to-English translation.

2.4 From breakthroughs to bottlenecks

The shift from statistical models to neural networks and LLMs, together with the transition from cascade to end-to-end systems, is driving major advancements in MI. Recent systems have significantly improved key performance metrics, including information fidelity, delivery fluency, latency reduction, and speaker fidelity and prosodic naturalness, often matching or surpassing human interpreter benchmarks (e.g., Cheng et al., 2024; Labiausse et al., 2025). By leveraging high-automation real-time processing, extensive multilingual knowledge bases, and continuous deep learning, current MI systems demonstrate significant potential to overcome the “cognitive resource and capacity limitations” (Lu, 2018) inherent in human SI – a task that demands the concurrent processing of auditory input and verbal output. Coupled with greater cost-effectiveness and enhanced operational reliability, these technological advancements have positioned MI systems as being increasingly competitive, as evidenced by their growing global adoption. These developments represent a paradigm shift with profound implications for interpreting practice, pedagogy, and research.

However, the nature of interpreting, which demands situated cognition and real-time processing, presents a far greater challenge for MI than for MT. Such complexity is increased further by the need to process spontaneous spoken language, which is highly context-dependent and often features colloquial expressions, irregularities, inaccuracies, and ellipses. Moreover, spoken language is rich in paralinguistic features such as tone, pace, pauses, and intonation, in addition to nonverbal cues such as facial expressions, gestures, and body movements. These complexities, combined with the design and operational limitations of MI systems, present persistent challenges (Lu, 2022, 2023), as indicated here:

- (1) multimodal processing of source prosody, speaker nonverbal cues, and supplementary materials (e.g., scripts, PPTs, audio, video);
- (2) challenges in source recognition arising from poor audio, noise, overlapping speech, accents, disfluencies, homophones, neologisms, proper names, culture-specific items, rare vocabulary, and irregular expressions;
- (3) a tendency towards literal translation rather than meaning-based translation, limiting the capture of pragmatic meaning, cultural and ideological nuances, emotional subtleties, and metaphorical language;
- (4) deficiencies in output naturalness, prosody, emotional resonance, and stylistic accuracy, compromising usability and communicative effectiveness;
- (5) underdeveloped multilingual systems having to deal with language pairs, directionality, and low-resource or minority languages;
- (6) limitations in employing adaptive strategies or tactics such as amplification, clarification, filtering, mitigation, and intercultural mediation;

Lu, X., & Fantinuoli, C. (2025). Machine and computer-assisted interpreting: Innovations in and implications for Interpreting practice, Pedagogy and Research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 1–22.

- (7) the inability to perform multiple roles and switch flexibly between consecutive, simultaneous, and whispered interpreting;
- (8) the lack of social identity, perspectives, emotions, and warmth, qualities that are intrinsic to human interpreters and essential in many contexts;
- (9) the incapacity to handle on-site crises such as interpersonal conflicts, disputes, and technical system failures effectively.

In addition, MI raises ethical concerns, including questions of accuracy, accountability, bias, and privacy, together with social risks such as cultural homogenization, the digital divide, and ideological implications that require further scrutiny. These and other unresolved challenges inherent in MI highlight the need for further research by both research communities (i.e., computational and interpreting), combining their distinct approaches and expertise to ensure the responsible and valuable deployment of MI.

3. Evolution of and state of the art in computer-assisted interpreting

Research on CAI has evolved significantly during the past decade. What began as a relatively niche area of enquiry, one pursued by a small circle of scholars, has grown into a dynamic and widely explored field. This expansion has been driven in part by the increasing availability and sophistication of AI-based technologies, which offer new opportunities for supporting interpreters before, during, and after their assignments. As both academic and professional interest intensifies, the scope of CAI research has broadened accordingly.

The current body of literature can be broadly categorized into three main areas. The first, and the most extensively studied, is SI, where real-time support tools have received sustained scholarly attention. The second concerns consecutive and dialogue interpreting – for example, that employed in healthcare and community settings – which remains comparatively underexplored, despite its practical significance. A third area, which is still emerging, investigates the ways in which digital technologies can be leveraged to enhance interpreter training, evaluating as it does both the pedagogical outcomes of adopting such technologies and the technological integration within educational frameworks.

3.1 Different areas of research

3.1.1 Simultaneous interpreting

Current research on CAI in SI highlights the significant strides being made in real-time lexical support. Ergonomically designed interfaces now enable interpreters to access terminology while performing with minimal distraction, responding as they have to previous concerns about cognitive overload. Empirical studies demonstrate measurable improvements in the

accuracy of interpreting so-called problem-triggers – notably, numbers, proper names, and specialized terms – when interpreters use ASR-enabled lookup systems (e.g., Defrancq et al., 2024; Li & Chmiel, 2024). These systems offer enhanced precision through deep neural-network-based ASR engines, particularly in high-resource languages (e.g., Fantinuoli, 2017; Pöchhacker, 2016).

AI advancements have refined CAI interfaces more. For instance, context-aware suggestion algorithms can now filter terminology dynamically based on discourse context and, as a result, reducing cognitive load and enhancing relevance (e.g., Vogler et al., 2019). Then, too, LLMs are now showing promise for personalized preparatory support. And auto-generating tailored glossaries and information workflows are being generated based on interpreter needs and prior usage patterns (e.g., Fantinuoli, 2023).

Nonetheless, challenges remain: ASR latency and overabundant suggestions can still hinder performance, and studies underscore the importance of balancing automation against interpreter autonomy (e.g., Defrancq & Fantinuoli, 2021; Russello & Carbutto, 2023). Overall, research in this area continues to push towards more responsive, intelligent, and contextually sensitive CAI tools as support for SI.

3.1.2 Consecutive and dialogue interpreting

Recent investigations into computer-assisted tools for consecutive interpreting have focused on integrating mobile devices into the interpreter’s workflow. Tablets and smartphones now support digital note-taking, on-the-fly terminology lookups, and contextual access to preparatory materials, as a result enhancing the ergonomics and efficiency in live settings (e.g., Goldsmith, 2018). Advanced prototypes are exploring hybrid workflows that incorporate ASR and MT: interpreters listen to a speaker, receive structured visual input (e.g., partial transcripts or term suggestions), and then render the interpreting outcome, in the process enhancing both their lexical precision and their recall (e.g., Chen & Kruger, 2023; Ünlü & Doğan, 2024).

Dialogue interpreting, including healthcare, legal, and community-based settings, remains understudied in CAI research, though. Its interactive multi-party nature presents challenges not yet widely resolved by current tools. The published literature lacks systematic empirical studies or design frameworks to support dialogic contexts, revealing a significant research gap and an important avenue for future enquiry and investigation (e.g. Tan et al., 2025).

3.1.3 Computer-assisted interpreter training

State-of-the-art research on computer-assisted interpreter training (CAIT) reveals a growing interest in incorporating digital tools into educational settings. CAIT platforms now include

Lu, X., & Fantinuoli, C. (2025). Machine and computer-assisted interpreting: Innovations in and implications for Interpreting practice, Pedagogy and Research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 1–22.

speech repositories, authoring systems, multimedia resources, and virtual environments that are tailored to autonomous and blended learning (e.g., Sandrelli & Hawkins, 2006).

Moreover, empirical studies demonstrate that CAIT enhances learner outcomes by boosting their motivation, reducing performance anxiety, and promoting self-paced practice (e.g., Chan, 2023). In addition, virtual reality applications and remote learning frameworks which offer immersive and flexible training that simulates professional interpreting scenarios are also being evaluated (e.g., Braun et al., 2020)

Emerging AI technologies are enhancing CAIT even more. Research has been exploring the use of LLMs to generate domain-specific speeches at appropriate levels of difficulty alongside text-to-speech engines that produce naturalistic audio input, offering scalable, customizable training materials (e.g., Fantinuoli, 2023). Moreover, the use of augmented reality has also been the object of enquiry (e.g., Gieshoff et al., 2024). However, challenges persist: technological infrastructure, the dearth of digital literacy among both trainers and students, and the management of cognitive load remain barriers. Researchers emphasize the need for pedagogical frameworks that support the holistic integration of CAIT tools, frameworks that are tailored to both individual and contextual factors, an area still in the early stages of development (e.g., Prandi, 2020).

3.2 Limitations of current research

3.2.1 User interfaces

The number of CAI tools explicitly designed to meet the needs of interpreters remains limited and, consequently, research on their design from a user interface or a user experience (UI/UX) perspective is equally scarce. This gap is significant, particularly in relation to human–machine interaction during the act of interpreting itself. Regardless of the modality, interpreting entails a substantial cognitive load and the introduction of software into the process represents an additional source of demand. Human-friendly interface design is therefore critical.

Despite this, little academic or publicly available research has examined UI/UX considerations in depth. Most studies either make no mention of interface design or present the design employed in their own system as a de facto standard. In many cases, this involves either a commercial application, most commonly InterpretBank (e.g., Prandi, 2023), or purpose-built interfaces created solely for experimental use (e.g., Tan et al., 2025). Notable exceptions exist where design – specifically, how information should be presented to interpreters to enhance usability – has been, at least in part, a focus. Examples include the EABM project (Defrancq & Fantinuoli, 2021) with its large-scale survey of preferences among professionals, Baselli’s (2023) work on the interface requirements for remote interpreting, and Frittella’s (2023) usability testing framework devised to achieve an interpreter-centred design.

3.2.2 Study participants

A striking characteristic of the existing CAI literature is its overwhelming reliance on interpreting students as study participants. While this choice is often pragmatic and justifiable, it has been repeatedly identified as a limitation (e.g., Defrancq & Fantinuoli, 2021). That said, studies involving experienced professional interpreters remain the exception rather than the rule, with only a handful of contributions specifically targeting this group (e.g., Bereznaya, 2022; Frittella, 2023). This imbalance inevitably raises questions about the transferability of research findings to real-world professional contexts in which workflows, cognitive strategies, and expectations may differ considerably from those of students. The lack of systematic engagement with professional interpreters also risks reinforcing their scepticism towards CAI tools, because opportunities to contribute feedback during tool design and evaluation are limited; moreover, concerns about the potential effects of such tools on interpreting processes remain insufficiently considered or responded to.

3.2.3 Reproducibility of studies

Another drawback of the current state of research is the difficulty of replicating or comparing studies. Most experiments use different tools, or different versions of the same tool, and often fail to report clearly on the technical performance of the tool itself. This makes replicability challenging and, even more critically, comparability between studies nearly impossible. While it is inevitable that tools will vary and evolve over time, it would be good practice for researchers to provide accompanying technical documentation for each study. Such reports should clearly describe the tool's design, including its user interface, and, where relevant, provide key performance indicators such as word error rates (in the case of ASR-based systems), latency measures, and transparent definitions of the way in which these metrics were obtained. Establishing shared reporting standards of this kind would allow for more meaningful comparisons to be made between studies and help to build a cumulative body of knowledge rather than a collection of isolated, methodologically incompatible findings.

4. Contributions in this issue

As previously noted, research on MI remains relatively scarce in the field of Translation Studies, whereas greater attention has been devoted to CAI. This special issue of *LANS-TTS*, which features the theme *Machine and Computer-Assisted Interpreting*, mirrors this trend. It comprises seven contributions, two of which focus on MI, while the remaining five explore CAI. Again in line with the previous trend, the two contributions on MI focus on quality evaluation, their aim being to enable either direct quality estimation in MI or a comparison of the product and the process between MI and human interpreting. The contributions on CAI explore various facets of the craft, including the assessment of CAI tools in remote SI, the implications of ASR

Lu, X., & Fantinuoli, C. (2025). Machine and computer-assisted interpreting: Innovations in and implications for Interpreting practice, Pedagogy and Research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 1–22.

for consecutive interpreting, the effect of multimodal inputs, and the interaction patterns involved in live captioning in SI.

Information fidelity is fundamental to the communicative effectiveness of interpreting and it has long been a central focus in Interpreting Studies. However, the potential application of Machine Translation Quality Estimation (MTQE) metrics, speech representation models, and LLMs to the automatic speech-based assessment of information fidelity in MI has yet to be fully explored or validated. The study conducted by **Xiaoman Wang** and **Binhua Wang** leveraged speech embeddings together with LLM to assess interpreting quality at the speech level. They also used ASR, LLM, and MTQE models such as COMET and TransQuest for minute-level text-based assessment. The results of their study underscore the potential of combining speech-to-speech and text-based metrics to advance the development of machine learning models for assessing interpreting quality. They propose a more comprehensive assessment framework that leverages the synergies between diverse types of assessment while taking into account the biases and limitations inherent in each individual approach. This proposed framework offers significant potential as an automated, cost-effective, and labour-efficient solution to estimating the quality of end-to-end MI systems directly.

Although various linguistic dimensions have been explored to differentiate human translation from MT, comparable investigations into human interpreting versus MI remain underexplored. In this regard, the study by **Yao Yao**, **Kanglong Liu**, **Kay Fan Andrew Cheung**, and **Dechao Li** explores syntactic complexity as a distinguishing factor between MI and human interpreting. Their objective in doing so is to advance our understanding of the differences between computational and human language processing mechanisms in interpreting. To this end, they employed machine learning classifiers to perform a multidimensional analysis of syntactic complexity, aiming to differentiate between interpreting by iFLYTEK, a cascade system, and that by professional interpreters, based on a comparable Chinese-to-English corpus drawn from government press conferences. They found that MI exhibited “additive complexity” driven by modular sequential processing within cascading architectures. In contrast, human interpreting demonstrated “integrative complexity” rooted in conceptually mediated processing that balances cognitive constraints against communicative goals. This study has the potential to offer an empirical foundation for developing more naturalistic MI systems that emulate professional interpreters, particularly their strengths in meaning-based and message-based interpreting.

Although several methodologies have been proposed in the past to evaluate CAI tools, the study by **Zhiqiang Du** and **Ricardo Muñoz Martín** offers a significant contribution in this regard by proposing a robust methodological framework for evaluating CAI tools in a specific setting: remote simultaneous interpreting (RSI). Moving beyond traditional information-processing models, these researchers adopted a situated cognition approach in order to assess the ways

in which interpreting trainees interact with digital glossaries in authentic multimodal environments. Through a mixed-methods, pretest–post-test design, the study compares the impact of using InterpretBank (experimental group) versus Microsoft Excel (control group) on term accuracy, speech fluency, cognitive effort, and overall performance. Their findings highlight the fact that while CAI tools can enhance interpreter output, their effectiveness is shaped by task complexity, individual adaptation strategies, and usability features such as error tolerance and search precision. This research also underscores the importance of involving interpreters in the design and evaluation of CAI tools and provides practical insights into tool development, interpreter training, and the integration of digital support in professional practice.

The study by **Xuejiao Peng, Xiangling Wang** and **Guangjiao Chen** advances our understanding of computer-assisted simultaneous interpreting (CASI) by examining how multimodal input – specifically the combination of audio, video, and textual elements – affects interpreting quality, cognitive load, and attention dynamics. Drawing on a mixed-methods approach that includes eye-tracking, quality assessment, and self-reported cognitive load measures, the research involved 30 student interpreters working from English into Chinese across four input conditions. The findings reveal that the richest multimodal condition (audio-video-text) resulted in both the highest interpreting quality and the lowest cognitive load; this suggests that well-designed CASI interfaces are able to facilitate cognitive processing during interpreting. Notably, the student interpreters prioritized textual input over video and audio cues, highlighting distinct strategies of attention coordination in multimodal environments. While the study acknowledges limitations related to participant expertise and experimental realism, it does offer valuable empirical evidence and insights for tool developers, trainers, and researchers in the evolving field of technology-mediated interpreting to take into consideration and to build on.

This study by **Zhibin Yu** and **Zhangminzi Shao** explored the effects of AI-powered automatic speech recognition technology on interpreter performance and cognitive load during consecutive interpreting, with particular attention to source speeches featuring varying accents. In a controlled experiment, the researchers compared the interpreting output across four tasks, both with and without ASR support, using both familiar and unfamiliar accents. A combination of objective and subjective metrics was used to assess the fidelity and fluency of the interpreting, the quality of the target language, and the cognitive load experienced by their cohort. The results indicate that ASR enhanced fidelity but negatively affected fluency, leading to slower delivery and more silent pauses. And whereas ASR had no significant effect on the overall quality or the cognitive load, accent familiarity influenced the way the interpreters engaged with the tool, an outcome that suggests a risk of over-reliance on ASR in more challenging scenarios. These findings underscore the nuanced role of ASR: while it can support

certain aspects of performance, its cognitive implications and interaction with speech variables require further empirical scrutiny.

The study in this collection by **Meng Guo, Yuxing Xie, Lili Han, Victoria Lai Cheng Lei** and **Defeng Li** investigates the integration of automatic live captioning into SI, offering new insights into the way interpreters engage with visual transcription in real time. Their research combines eye-tracking data, temporal metrics, and user feedback to examine the manner in which live captioning affects visual attention, temporal synchronization, and cognitive processing. The findings reveal a marked reliance on visual transcription, with captioning becoming a dominant support modality during interpreting. This reliance influenced the interpreters' timing and information management, as is evidenced by the increase in ear–voice span (EVS) over time. While the participants generally perceived the live captioning as being helpful, they also expressed the need for more customizable and adaptable interfaces.

Finally, **Wenchao Su** and **Defeng Li** present an eye-tracking investigation into the ways in which professional interpreters apportion visual attention while working with real-time ASR-generated source captions and AST-generated target subtitles during technology-assisted SI. Their study examines both the L1–L2 and the L2–L1 directions, revealing that ASR captions generally attract more attention than AST subtitles, particularly in the L1–L2 mode; and that those captions and subtitles displayed in the interpreters' native language capture more gaze time. By disentangling the interplay of bottom-up and top-down factors in attention allocation, the study provides empirical insights that can inform targeted training for interpreters in technology-rich environments.

5. Final remarks

The emergence of MI and the continuing advancements in CAI are serving to redefine the global interpreting ecosystem. This special issue offers valuable contributions to the growing discourse on the transformative role of MI and CAI in shaping interpreting practice, pedagogy, and research.

A major challenge in MI lies in fostering collaboration between the computational and the interpreting communities so as to enhance systems and workflows, focusing on process and capability comparisons, leveraging interpreters' strengths, developing high-quality corpora, and refining evaluation methods. Another critical priority is developing adaptable MI systems that accommodate diverse contexts (e.g., conference types, domains, speech types) and source-language features (e.g., accents, information density, linguistic features, and delivery styles), while supporting complex modes such as simultaneous–consecutive transitions. It is essential to explore the pathways and mechanisms for human–machine collaboration across varying scenarios, source-language conditions, and interpreting modes to establish a robust

human–machine hybrid model. Moreover, it is crucial to reassess the ways in which MI reshapes interpreting constructs, competence, processes, products, effects, and its broader implications for education, professional practice, and ethics.

As current research on CAI continues to demonstrate the potential benefits of real-time tools, particularly in the context of conference interpreting, it is important to recognize the limitations of this narrow focus. The majority of studies concentrate on high-level multilingual events which represent only a small fraction of real-world interpreting practice. Far less attention has been paid to other settings such as healthcare, legal, and community interpreting, where CAI tools may hold the potential (or reveal limitations) for improving working conditions, accessibility, and quality. Moreover, it remains largely unexplored whether such tools could help to expand the pool of available interpreters in community settings, especially in countries that are facing workforce shortages. Another significant limitation is the dominance of widely spoken and AI-favoured language pairs in experimental research: it is still unclear, for instance, how well CAI technologies perform with, or could be adapted to, under-resourced languages.

A further constraint lies in the widespread reliance on student interpreters as research participants. While this practice is understandable from a logistical standpoint, it raises questions about the generalizability of findings to professional settings. Interpreting students differ significantly from experienced practitioners in terms of their cognitive strategies, domain expertise, and tool-usage habits. To build a more robust evidence base, it would be highly beneficial for large institutions, such as international organizations, government bodies, and professional associations, to invest in systematic studies that involve trained working interpreters across all modalities. The scope for future research is vast and it is imperative that the scholarly community broaden its lens both in interpreting contexts and in language diversity while also envisaging novel applications that could support the profession more inclusively and effectively.

Lu, X., & Fantinuoli, C. (2025). Machine and computer-assisted interpreting: Innovations in and implications for Interpreting practice, Pedagogy and Research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 1–22.

Funding

The research is supported by the National Social Science Fund of China “Information Processing Routes and Mechanisms in Chinese-English Simultaneous Interpreting: a Corpus-based Study” [Project Number: 22AYY005].

References

- Baselli, V. (2023). Developing a new CAI tool for RSI interpreters’ training: A pilot study. *Proceedings of the International Conference HiT-IT 2023*, 157–166. https://doi.org/10.26615/issn.2683-0078.2023_014
- Braun, S., Davitti, E., & Slater, C. (2020). ‘It’s like being in bubbles’: Affordances and challenges of virtual learning environments for collaborative learning in interpreter education. *The Interpreter and Translator Trainer*, 14(3), 259–278. <https://doi.org/10.1080/1750399X.2020.1800362>
- Bereznaya, V. (2022). *Can computer-assisted interpreting (CAI) tools be reliable artificial boothmates?: Testing the usability of InterpretBank’s automatic speech recognition feature in a remote interpreting setting* [Unpublished master’s thesis]. University of Vienna. <https://doi.org/10.25365/thesis.71842>
- Chan, V. (2023). Research on computer-assisted interpreter training: A review of studies from 2013 to 2023. *SN Computer Science*, 4, Article 648. <https://doi.org/10.1007/s42979-023-02072-w>
- Chen, M., Duquenne, P.-A., Andrews, P., Kao, J., Mourachko, A., Schwenk, H., & Costa-jussà, M. R. (2022). BLASER: A text-free speech-to-speech translation evaluation metric. ArXiv. <https://doi.org/10.48550/arXiv.2212.08486>
- Chen, S., Kruger, J.-L. (2023). The effectiveness of computer-assisted interpreting: A preliminary study based on English–Chinese consecutive interpreting. *Translation and Interpreting Studies*, 18(3), 399–420. <https://doi.org/10.1075/tis.21036.che>
- Chen, X., Zhang, S., Bai, Q., Chen, K., & Nakamura, S. (2024). LLaST: Improved end-to-end speech translation system leveraged by large language models. ArXiv. <https://doi.org/10.18653/v1/2024.findings-acl.416>
- Cheng, S., Huang, Z., Ko, T., Li, H., Peng, N., Xu, L., & Q. Zhang (2024). Towards achieving human parity on end-to-end simultaneous speech translation via LLM agent. ArXiv. <https://doi.org/10.48550/arXiv.2407.21646>
- Cho, E., Fugen, C., Herrmann, T., Kilgour, K., Mediani, M., Mohr, C., Niehues, J., Rottmann, K., Saam, C., Stucker, S., & Waibel, A. (2013). A real-world system for simultaneous translation of German lectures. *Interspeech 2013*, 3473–3477. <https://doi.org/10.21437/Interspeech.2013-612>
- Cho, K., & Esipova, M. (2016). Can neural machine translation do simultaneous translation? ArXiv. <https://doi.org/10.48550/arXiv.1606.02012>
- Defrancq, B., & Fantinuoli, C. (2021). Automatic speech recognition in the booth: Assessment of system performance, interpreters’ performances and interactions in the context of numbers. *Target*,

Lu, X., & Fantinuoli, C. (2025). Machine and computer-assisted interpreting: Innovations in and implications for Interpreting practice, Pedagogy and Research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 1–22.

International Journal of Translation Studies, 33(1), 77–102. <https://doi.org/10.1075/target.19166.def>

- Defrancq, B., Snoeck, H., & Fantinuoli, C. (2024). Interpreters' performances and cognitive load in the context of a CAI tool. In M. Winters, S. Deane-Cox, & U. Böser (Eds.), *Translation, interpreting and technological change. Innovations in research, practice and training* (pp. 38–58). Bloomsbury. <https://doi.org/10.5040/9781350212978.0009>
- Dong, Q., Huang, Z., Tian, Q., Xu, C., Ko, T., Zhao, Y., Feng, S., Li, T., Wang, K., Cheng, X., Yue, F., Bai, Y., Chen, X., Lu, L., Ma, Z., Wang, Y., Wang, M., & Wang, Y. (2023). PolyVoice: Language models for speech to speech translation. ArXiv. <https://doi.org/10.48550/arXiv.2306.02982>
- Dua, S., Kumar, S. S., Albagory, Y., Ramalingam, R., Dumka, A., Singh, R., Rashid, M., Gehlot, A., Sultan S. Alshamrani, S. S., & AlGhamdi, A. S. (2022). Developing a speech recognition system for recognizing tonal speech signals using a convolutional neural network. *Applied Sciences*, 12, Article 6223. <https://doi.org/10.3390/app12126223>
- Fang, Q., & Feng, Y. (2023). Understanding and bridging the modality gap for speech translation. ArXiv. <https://doi.org/10.48550/arXiv.2305.08706>
- Fang, Q., Zhang, S., Ma, Z., Zhang, M., & F. Yang (2024). Can we achieve high-quality direct speech-to-speech translation without parallel speech data? ArXiv. <https://doi.org/10.48550/arXiv.2406.07289>
- Fantinuoli, C. (2017). Speech recognition in the interpreter workstation. *Proceedings of the Translating and the Computer*, 39, 25–34.
- Fantinuoli, C. (2023). Towards AI-enhanced computer-assisted interpreting. In G. Corpas Pastor & B. Defrancq (Eds.), *IVITRA research in linguistics and literature* (pp. 46–71). John Benjamins. <https://doi.org/10.1075/ivittra.37.03fan>
- Fantinuoli, C. (2018). Interpreting and technology: The upcoming technological turn. In C. Fantinuoli (Ed.), *Interpreting and technology* (pp. 1–12). Language Science Press. <https://doi.org/10.5281/zenodo.1493289>
- Fantinuoli, C., & Prandi, B. (2021). Towards the evaluation of automatic simultaneous speech translation from a communicative perspective. *Proceedings of the 18th IWSLT*, 245–254. <https://doi.org/10.18653/v1/2021.iwslt-1.29>
- Frittella, F. M. (2023). *Usability research for interpreter-centred technology: The case study of SmarTerp*. Language Science Press. <https://doi.org/10.5281/ZENODO.7376351>
- Fügen, C., Waibel, A., & Kolss, M. (2007). Simultaneous translation of lectures and speeches. *Machine Translation*, 21, 209–252. <https://doi.org/10.1007/s10590-008-9047-0>
- Gao, Y., Zhou, B., Gu, L., Sarikaya, R., Afify, M., Kuo, H., Zhu, W., Deng, Y., Prosser, C., Zhang, W., & Besacier, L. (2006). IBM MASTOR: Multilingual automatic speech-to-speech translator. *MST '06: Proceedings of the Workshop on Medical Speech Translation*, 53–56, Association for Computational Linguistics. <https://doi.org/10.3115/1706257.1706268>

- Lu, X., & Fantinuoli, C. (2025). Machine and computer-assisted interpreting: Innovations in and implications for Interpreting practice, Pedagogy and Research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 1–22.
- Gieshoff, A. C., Schuler, M., & Zaniyar J. (2024). The augmented interpreter: An exploratory study of the usability of augmented reality technology in interpreting. *Interpreting: International Journal of Research and Practice in Interpreting*, 26(2), 282–315. <https://doi.org/10.1075/intp.00108.gie>
- Glocknitzer, J. (2020). *Arzt–Patient-Kommunikation mit SayHi Translate* [Unpublished master’s thesis]. University of Vienna.
- Goldsmith, J. (2018). Tablet interpreting: Consecutive interpreting 2.0. *Translation and Interpreting Studies*, 13(3), 342–365. <https://doi.org/10.1075/tis.00020.gol>
- Grissom II, A. C., Boyd-Graber, J., He, H., Morgan, J., & Daumé III, H. (2014). Don’t until the final verb wait: Reinforcement learning for simultaneous machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1342–1352. <https://doi.org/10.3115/v1/D14-1140>
- Hamon, O., Fügen, D., Mostefa, D., Arranz, V., Kolss, M., Waibel, A., & Choukri, K. (2009). End-to-end evaluation in simultaneous translation. *EACL’09: Proceedings of the 12th Conference of the European Chapter of the ACL, Association for Computational Linguistics*, 345–353. <https://doi.org/10.3115/1609067.1609105>
- Huang, W. C., Peloquin, B., Kao, J., Wang, C., Gong, H., Salesky, E., & Chen, P. J. (2023). A holistic cascade system, benchmark, and human evaluation protocol for expressive speech-to-speech translation. *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 1–5. <https://doi.org/10.1109/ICASSP49357.2023.10096183>
- Huang, Z., Ye, R., Ko, T., Dong, Q., Cheng, S., Wang, M., & Li, H. (2023). Speech translation with large language models: An industrial practice. ArXiv. <https://doi.org/10.48550/arXiv.2312.13585>
- Jekat, S. J., & Klein, A. (1996). Machine interpretation open problems and some solutions. *Interpreting: International Journal of Research and Practice in Interpreting*, 1(1), 7–20. <https://doi.org/10.1075/intp.1.1.02jek>
- Jia, Y., Weiss, R. J., Biadys, F., Macherey, W., Johnson, M., Chen, Z., & Wu, Y. (2019). Direct speech-to-speech translation with a sequence-to-sequence model. ArXiv. <https://doi.org/10.48550/arXiv.1904.06037>
- Kikui, G. I., Yamamoto, S., Takezawa, T., & Sumita, E. (2006). Comparative study on corpora for speech translation. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5), 1674–1682. <https://doi.org/10.1109/TASL.2006.878262>
- Labiausse, T., Mazaré, L., Grave, E., Pérez, P., Défossez, A., & Zeghidour, N. (2025). High-fidelity simultaneous speech-to-speech translation. ArXiv. <https://doi.org/10.48550/arXiv.2502.03382>
- Lam, T. K., Schamoni, S., & Riezler, S. (2023). Make more of your data: Minimal effort data augmentation for automatic speech recognition and translation. *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 1–5. <https://doi.org/10.1109/ICASSP49357.2023.10094564>
- Le N., Lecouteux, B., & Besacier, L. (2018). Automatic quality estimation for speech translation using joint ASR and MT features. *Machine Translation*, 32(4), 325–351. <https://doi.org/10.1007/s10590-018-9218-6>

- Lu, X., & Fantinuoli, C. (2025). Machine and computer-assisted interpreting: Innovations in and implications for Interpreting practice, Pedagogy and Research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 1–22.
- Li, T., & Chmiel, A. (2024). Automatic subtitles increase accuracy and decrease cognitive load in simultaneous interpreting. *Interpreting: International Journal of Research and Practice in Interpreting*, 26(2), 253–281. <https://doi.org/10.1075/intp.00111.li>
- Liu, Y., & Liang, J. (2024). Multidimensional comparison of Chinese-English interpreting outputs from human and machine: Implications for interpreting education in the machine-translation age. *Linguistics and Education*, 80, Article 101273. <https://doi.org/10.1016/j.linged.2024.101273>
- Lu, X. (2018). Propositional information loss in English-to-Chinese simultaneous conference interpreting: A corpus-based study. *Babel*, 64(5–6), 792–818. <https://doi.org/10.1075/babel.00070.lu>
- Lu, X. (2022). Comparing the quality and processes of Chinese–English simultaneous interpreting by interpreters and a machine. *Foreign Language Teaching and Research*, 54(4), 600–610.
- Lu, X. (2023). Simultaneous interpreting by interpreters and machine: Cognitive processes, competences, qualities and future trends. *Chinese Translators Journal*, 3, 135–141.
- Luperfoy, S. (1996). Machine interpretation of bilingual dialogue. *Interpreting: International Journal of Research and Practice in Interpreting*, 1(2), 213–233. <https://doi.org/10.1075/intp.1.2.03lup>
- Ma, M., Huang, L., Xiong, H., Liu, K., Zhang, C., He, Z., Liu, H., Li, X., & Wang, H. (2019). Stacl: Simultaneous translation with integrated anticipation and controllable latency. ArXiv. <https://doi.org/10.48550/arXiv.1810.08398>
- Ma, Z., Fang, Q., Zhang, S., Guo, S., Feng, Y., & Zhang, M. (2024). A non-autoregressive generation framework for end-to-end simultaneous speech-to-any translation. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 1557–1575. <https://doi.org/10.18653/v1/2024.acl-long.85>
- Mi, C., Xie, L., & Zhang, Y. (2022). Improving data augmentation for low resource speech-to-text translation with diverse paraphrasing. *Neural Networks*, 148, 194–205. <https://doi.org/10.1016/j.neunet.2022.01.016>
- Mujadia, V., Mishra, P., & Sharma, D. M. (2025). Disfluency processing for cascaded speech translation involving English and Indian languages. *Lang Resources & Evaluation*, 59, 2653–2686. <https://doi.org/10.1007/s10579-025-09818-3>
- Murata, M., Ohno, T., Matsubara, S., & Inagaki, Y. (2010). Construction of chunk-aligned bilingual lecture corpus for simultaneous machine translation. *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010*. European Language Resources Association.
- Nakamura, S. (2009). Overcoming the language barrier with speech translation technology. *Quarterly Review*, 31, 35–48.
- Papi, S., Gaido, M., Karakanta, A., Cettolo, M., Negri, M., & Turchi, M. (2023). Direct speech translation for automatic subtitling. *Transactions of the Association for Computational Linguistics*, 11, 1355–1376. https://doi.org/10.1162/tacl_a_00607
- Pöschhacker, F. (2016). *Introducing interpreting studies* (2nd ed). Routledge. <https://doi.org/10.4324/9781315649573>

Lu, X., & Fantinuoli, C. (2025). Machine and computer-assisted interpreting: Innovations in and implications for Interpreting practice, Pedagogy and Research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 1–22.

Pöschhacker, F. (2024). Is machine interpreting interpreting? *Translation Spaces*, 1–21 <https://doi.org/10.1075/ts.23028.poc>

Prandi, B. (2020). The use of CAI tools in interpreter training: Where are we now and where do we go from here? *inTRAlinea Special Issue*, 1–10.

Prandi, B. (2023). *Computer-assisted simultaneous interpreting: A cognitive–experimental study on terminology*. Language Science Press.

Rodríguez González, E., Saeed M., Korybski, T., Davitti, E., Braun S. (2023). *Assessing the impact of automatic speech recognition on remote simultaneous interpreting performance using the NTR Model. Proceedings of the International Workshop on Interpreting Technologies SAY IT Again 2023*, 1–8.

Russello, C., & Carbutto, M. (2023). Enhancing numerical accuracy in simultaneous interpreting: A comparative study of human and AI-based support. *The Interpreters' Newsletter*, 28, 69–90. <https://doi.org/10.13137/2421-714X/35551>

Sakamoto, A., Abe, K., Sumita, K., & Kamatani, S. (2013). Evaluation of a simultaneous interpretation system and analysis of speech log for user experience assessment. *Proceedings of the 10th International Workshop on Spoken Language Translation: Papers*.

Salevsky, H. (1993). The distinctive nature of interpreting studies. *Target*, 5(2), 149–167. <https://doi.org/10.1075/target.5.2.03sal>

Sandrelli, A., & Hawkins, J. (2006). Computer Assisted Interpreter Training (CAIT): What is the way forward? *Proceedings of Accessible Technologies in Translation and Interpreting Conference*.

SEAMLESS Communication Team (2025). Joint speech and text machine translation for up to 100 languages. *Nature*, 637, 587–593. <https://doi.org/10.1038/s41586-024-08359-z>

Shimizu, H., Neubig, G., Sakti, S., Toda, T., & Nakamura, S. (2013). Constructing a speech translation system using simultaneous interpretation data. *Proceedings of the 10th International Workshop on Spoken Language Translation: Papers*.

Shin, J., Georgiou, P. G., & Narayanan, S. (2013). Enabling effective design of multimodal interfaces for speech-to-speech translation system: An empirical study of longitudinal user behaviors over time and user strategies for coping with errors. *Computer Speech & Language*, 27(2), 554–571. <https://doi.org/10.1016/j.csl.2012.02.001>

Siahbani, M., Shavarani, H.S., Alinejad, A., & Sarkar, A. (2018). Simultaneous translation using optimized segmentation. *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas*, 154–167. Association for Machine Translation in the Americas.

Sridhar, V. K. R., Bangalore, S., & Narayanan, S. (2013). Enriching machine-mediated speech-to-speech translation using contextual information. *Computer Speech & Language*, 27(2), 554–571. <https://doi.org/10.1016/j.csl.2011.08.001>

Stede, M., & Schmitz, B. (2000). Discourse particles and discourse functions. *Machine Translation*, 15, 125–147. <https://doi.org/10.1023/A:101111203187>

- Lu, X., & Fantinuoli, C. (2025). Machine and computer-assisted interpreting: Innovations in and implications for Interpreting practice, Pedagogy and Research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 24, 1–22.
- Székely, É., Steiner, I., Ahmed, Z., & Carson-Berndsen, J. (2014). Facial expression-based affective speech translation. *Journal on Multimodal User Interfaces*, 8, 87–96. <https://doi.org/10.1007/s12193-013-0128-x>
- Tan, S., Orăsan, C., & Braun, S. (2025). Integrating automatic speech recognition into remote healthcare interpreting: A pilot study of its impact on interpreting quality. ArXiv. <https://doi.org/10.48550/arXiv.2502.03381>
- Tripepi Winteringham, S. (2010). The usefulness of ICTs in interpreting practice. *The Interpreters' Newsletter*, 15, 87–99.
- Ünlü, C., Doğan, A. (2024). Enhancing consecutive interpreting with ASR: Sight-Terp as a computer-assisted interpreting tool. *Revista Tradumàtica. Technologies de la Traducció*, 22, 401–425. <https://doi.org/10.5565/rev/tradumatica.382>
- Vogler, N., Stewart, C., & Neubig, G. (2019). Lost in interpretation: Predicting untranslated terminology in simultaneous interpretation. ArXiv. <https://doi.org/10.48550/arXiv.1904.00930>
- Wang, X. L., Finch, A., Utiyama, M., & Sumita, E. (2016). An efficient and effective online sentence segmenter for simultaneous interpretation. *Proceedings of the 3rd Workshop on Asian Translation*, The COLING 2016 Organizing Committee, 139–148.
- Wonisch, A. (2017). *Skype Translator: Funktionsweise und Analyse der Dolmetschleistung in der Sprachrichtung Englisch–Deutsch* [Unpublished master's thesis]. University of Vienna.
- Wu, Z., Caglayan, O., Ive, J., Wang, J., & Specia, L. (2019). Transformer-based cascaded multimodal speech translation. *Proceedings of the 16th International Conference on Spoken Language Translation*. Association for Computational Linguistics.